# APPENDIX B

# LIST OF EXPERTS AND OBSERVERS

# Workshop on Selecting Input Distributions for Probabilistic Assessment

U.S. Environmental Protection Agency
New York, NY
April 21-22, 1998

# List of Experts

**Sheila Abraham**
Environmental Specialist
Risk Assessment/Management
Northeast District Office
Ohio Environmental
Protection Agency
2110 East Aurora Road
Twinsburg, OH 44087
330-963-1290
Fax: 330-487-0769
E-mail:sabraham@epa.state.oh.us

**Hans Allender**
U.S. Environmental
Protection Agency
401 M Street, SW (7509)
Washington, DC 20460
703-305-7883
E-mail: allender.hans@
epamail.epa.gov

**Timothy Barry**
Office of Science Policy,
Planning, and Evaluation
U.S. Environmental
Protection Agency
401 M Street, SW (2174)
Washington, DC 20460
202-260-2038
E-mail: barry.timothy@
epamail.epa.gov

**Robert Blaisdell**
Associate Toxicologist
California Office of Environmental
Health Hazard Assessment
2151 Berkeley Way
Annex 11 - 2nd Floor
Berkeley, CA 94704
510-540-3487
Fax: 510-540-2923
E-mail: bblaisde@
berkeley.cahwnet.gov

**David Burmaster**
President
Alceon Corporation
P.O. Box 382669
Cambridge, MA 02238-2669
617-864-4300
Fax: 617-864-9954
E-mail: deb@alceon.com

**Christopher Frey**
Assistant Professor
Department of Civil Engineering
North Carolina State University
P.O. Box 7908
Raleigh, NC 27695-7908
919-515-1155
Fax: 919-515-7908
E-mail: frey@eos.ncsu.edu

**Susan Griffin**
Environmental Scientist
Superfund Remedial Branch
Hazardous Waste
Management Division
U.S. Environmental
Protection Agency
999 18th Street (8EPR-PS)
Suite 500
Denver, CO 80202-2466
303-312-6651
Fax: 303-312-6065
E-mail: griffin.susan@
epamail.epa.gov

**Bruce Hope**
Environmental Toxicologist
Oregon Department of
Environmental Quality
811 Southwest 6th Avenue
Portland, OR 97204
503-229-6251
Fax: 503-229-6977
E-mail: hope.bruce@deq.state.or.us

**William Huber**
President
Quantitative Decisions
539 Valley View Road
Merion, PA 19066
610-771-0606
Fax: 610-771-0607
E-mail: whuber@quantdec.com

**Robert Lee**
Risk Analyst
Golder Associates, Inc.
4104 148th Avenue, NW
Redmond, WA 98052
206-367-2673
Fax: 206-616-4875
E-mail: rclee@u.washington.edu

**David Miller**
Chemist
Office of Pesticide Programs
Health Effects Division
U.S. Environmental
Protection Agency
401 M Street, SW (7509)
Washington, DC 20460
703-305-5352
Fax: 703-305-5147
E-mail: miller.david@
epamail.epa.gov

**Samuel Morris**
Environmental Scientist
Deputy Division Head
Brookhaven National Laboratory
Building 815
815 Rutherford Avenue
Upton, NY 11973
516-344-2018
Fax: 516-344-7905
E-mail: morris3@bnl.gov

**Jacqueline Moya**
Environmental Engineer
National Center for
Environmental Assessment
Office of Research and Development
U.S. Environmental
Protection Agency
401 M Street, SW (8623D)
Washington, DC 20460
202-564-3245
Fax: 202-565-0052
E-mail: moya.jacqueline@
epamail.epa.gov

**Christopher Portier**
Chief
Laboratory of Computational
Biology and Risk Analysis
National Institute of
Environmental Health Sciences
P.O. Box 12233 (MD- A306)
Research Triangle Park, NC 27709
919-541-4999
Fax: 919-541-1479
E-mail: portier@niehs.nih.gov

**P. Barry Ryan**
Professor
Exposure Assessment and
Environmental Chemistry
Rollins School of Public Health
Emory University
1518 Clifton Road, NE
Atlanta, GA 30322
404-727-3826
Fax: 404-727-8744
E-mail: bryan@sph.emory.edu

**Brian Sassaman**
Bioenvironmental Engineer
U.S. Air Force
DET1, HSC/OEMH
2402 E Drive
Brooks Air Force Base, TX 78235-5114
210-536-6122
Fax: 210-536-1130
E-mail: brian.sassaman@
guardian.brooks.af.mil

**Ted Simon**
Toxicologist
Federal Facilities Branch
Waste Management Division
U.S. Environmental Protection Agency
Atlanta Federal Center
61 Forsyth Street, SW
Atlanta, GA 30303-3415
404-562-8642
Fax: 404-562-8566
E-mail: simon.ted@epamail.epa.gov

**Mitchell J. Small**
Professor
Departments of Civil & Environmental
Engineering and Engineering
& Public Policy
Carnegie Mellon University
Porter Hall 119, Frew Street
Pittsburgh, PA 15213-3890
412-268-8782
Fax: 412-268-7813
E-mail: ms35@andrew.cmu.edu

**Edward Stanek**
Professor of Biostatistics
Department of Biostatistics
and Epidemiology
University of Massachusetts
404 Arnold Hall
Amherst, MA 01003-0430
413-545-4603
Fax: 413-545-1645
E-mail: stanek@schoolph.umass.edu

**Alan Stern**
Acting Chief
Bureau of Risk Analysis
Division of Science and Research
New Jersey Department of
Environmental Protection
401 East State Street
P.O. Box 409
Trenton, NJ 08625
609-633-2374
Fax: 609-292-7340
E-mail: astern@dep.state.nj.us

**Paul White**
Environmental Engineer
National Center for
Environmental Assessment
Office of Research and Development
U.S. Environmental
Protection Agency
401 M Street, SW (8623D)
Washington, DC 20460
202-564-3289
Fax: 202-565-0078
E-mail: white.paul@epamail.epa.gov

(over)

# Workshop on Selecting Input Distributions for Probabilistic Assessment

U.S. Environmental Protection Agency
New York, NY
April 21-22, 1998

# Final List of Observers

**Samantha Bates**
Graduate Student/
Research Assistant
Department of Statistics
University of Washington
Box 354322
Seattle, WA 98195
206-543-8484
Fax: 206-685-7419
E-mail: sam@stat.washington.edu

**Steve Chang**
Environmental Engineer
Office of Emergency and
Remedial Response
U.S. Environmental
Protection Agency
401 M Street, SW (5204G)
Washington, DC 20460
703-603-9017
Fax: 703-603-9103
E-mail: chang.steve@
epamail.epa.gov

**Helen Chernoff**
Senior Scientist
TAMS Consultants, Inc.
655 Third Avenue
New York, NY 10017
212-867-1777
Fax: 212-697-6354
E-mail: hchernoff@
tamsconsultants.com

**Christine Daily**
Health Physicist
Radiation Protection &
Health Effects Branch
Division of Regulatory Applications
U.S. Nuclear Regulatory Commission
(T-9C24)
Washington, DC 20555
301-415-6026
Fax: 301-415-5385
E-mail: cxd@nrc.gov

**Emran Dawoud**
Human Health Risk Assessor
Toxicology and Risk Analysis Section
Life Science Division
Oak Ridge National Laboratory
1060 Commerce Park Drive (MS-6480)
Oak Ridge, TN 37830
423-241-4739
Fax: 423-574-0004
E-mail: dawoudea@ornl.gov

**Audrey Galizia**
Environmental Scientist
Program Support Branch
Emergency and Remedial
Response Division
U.S. Environmental Protection Agency
290 Broadway
New York, NY 10007
212-637-4352
Fax: 212-637-4360
E-mail: galizia.audrey@
epamail.epa.gov

**Ed Garvey**
TAMS Consultants
300 Broadacres Drive
Bloomfield, NJ 07003
973-338-6680
Fax: 973-338-1052
E-mail: egarvey@
tamsconsultants.com

**Gerry Harris**
UMDNJ-RWJMS
UMDNJ-EOSHI
Rutgers University
170 Frelinghuysen Road - Room 234
Piscataway, NJ 08855-1179
732-235-5069
E-mail: gharris@gpph.rutgers.edu

**David Hohreiter**
Senior Scientist
BBL
6723 Towpath Road
P.O. Box 66
Syracuse, NY 13214
315-446-9120
Fax: 315-446-7485
E-mail: dh%bbl@mcimail.com

mental Scientist
Hazardous Waste
Management Division
U.S. Environmental
Protection Agency
841 Chestnut Building (3H541)
Philadelphia, PA 19107
215-566-3324
E-mail: jafolla.nancy@
epamail.epa.gov

**Alan Kao**
Senior Science Advisor
ENVIRON Corporation
4350 North Fairfax Drive - Suite 300
Arlington, VA 22203
703-516-2308
Fax: 703-516-2393

**Steve Knott**
Executive Director, Risk
Assessment Forum
Office of Research and Development
National Center for
Environmental Assessment
U.S. Environmental
Protection Agency
401 M Street, SW (8601-D)
Washington, DC 20460
202-564-3359
Fax: 202-565-0062
E-mail: knott.steve@epamail.epa.gov

**Stephen Kroner**
Environmental Scientist
Office of Solid Waste
U.S. Environmental
Protection Agency
401 M Street, SW (5307W)
Washington, DC 20460
703-308-0468
E-mail: kroner.stephen@
epamail.epa.gov

**Anne LeHuray**
Regional Risk Assessment Lead
Foster-Wheeler
Environmental Corporation
8100 Professional Place - Suite 308
Lanham, MD 20785
301-429-2116
Fax: 301-429-2111
E-mail: alehuray@fwenc.com

**Toby Levin**
Attorney
Advertising Practices
Federal Trade Commission
601 Pennsylvania Avenue, NW
Suite 4110
Washington, DC 20852
202-326-3156
Fax: 202-326-3259

**Lawrence Myers**
Statistician
Research Triangle Institute
P.O. Box 12194
Research Triangle Park, NC 27709
919-541-6932
Fax: 919-541-5966
E-mail: lem@rti.org

**Marian Olsen**
Environmental Scientist
Technical Support Section
Program Support Branch
Emergency and Remedial
Response Division
U.S. Environmental Protection Agency
290 Broadway
New York, NY 10007
212-637-4313
Fax: 212-637-4360
E-mail: olsen.marion@epamail.epa.gov

**Lenwood Owens**
President
Boiler Servicing
1 Laguardia Road
Chester, NY 10918

**Zubair Saleem**
Office of Solid Waste
U.S. Environmental Protection Agency
401 M Street, SW (5307W)
Washington, DC 20460
703-308-0467
Fax: 703-308-0511
E-mail: saleem.zubair@
epamail.epa.gov

**Swati Tappin**
Research Scientist
New Jersey Department of
Environmental Protection
401 East State Street
P.O. Box 413
Trenton, NJ 08625
609-633-1348

**Joan Tell**
Senior Environmental Scientist
Exxon Biomedical
Mettlers Road (CN 2350)
East Millston, NJ 08875
732-873-6304
Fax: 732-873-6009
E-mail: joan.tell@exxon.sprint.com

**Bill Wood**
Executive Director, Risk
Assessment Forum
Office of Research and Development
National Center for
Environmental Assessment
U.S. Environmental
Protection Agency
401 M Street, SW (8601-D)
Washington, DC 20460
202-564-3358
Fax: 202-565-0062
E-mail: wood.bill@epamail.epa.gov

# APPENDIX C

# AGENDA

United States
Environmental Protection Agency
Risk Assessment Forum

# Workshop on Selecting Input Distributions for Probabilistic Assessment

U.S. Environmental Protection Agency
New York, NY
April 21-22, 1998

# Agenda

**Workshop Chair:** **Christopher Frey**
**North Carolina State University**

**TUESDAY, APRIL 21, 1998**

8:00AM **Registration/Check-In**

9:00AM **Welcome Remarks**
*Representative from Region 2, U.S. Environmental Protection Agency (U.S. EPA), New York, NY*

9:10AM **Overview and Background**
*Steve Knott, U.S. EPA, Office of Research and Development (ORD), Risk Assessment Forum, Washington, DC*

9:30AM **Workshop Structure and Objectives**
*Christopher Frey, Workshop Chair*

9:45AM **Introduction of Invited Experts**

10:00AM **Presentation: Issue Paper #1 - Evaluating Representativeness of Exposure Factors Data**
*Jacqueline Moya, U.S. EPA, National Center for Environmental Assessment (NCEA), Washington, DC*

10:15AM **Presentation: Issue Paper #2 - Empirical Distribution Functions and Non-Parametric Simulation**
*Tim Barry, U.S. EPA, NCEA, Washington, DC*

10:30AM B R E A K

10:45AM **Charge to the Panel**
*Christopher Frey, Workshop Chair*

11:00AM **Discussion on Issue #1: Representativeness**

12:00PM L U N C H

(over)

## TUESDAY, APRIL 21, 1998 (continued)

1:30PM **Discussion on Issue #1 Continues**

3:00PM B R E A K

3:15PM **Discussion on Issue #1 Continues**
*Christopher Frey, Workshop Chair*

4:15PM **Observer Comments**

4:45PM **Review of Charge for Day Two**
*Christopher Frey, Workshop Chair*

- Writing Assignments

5:00PM A D J O U R N

## WEDNESDAY, APRIL 22, 1998

8:30AM **Planning and Logistics**
*Christopher Frey, Workshop Chair*

8:40AM **Summary of Discussion on Issue #1**

10:00AM B R E A K

10:15AM **Discussion on Issue #2**: **Empirical Distribution Functions and Resembling Versus Parametric Distributions**

12:00PM L U N C H

1:30PM **Discussion on Issue #2 Continues**

3:00PM B R E A K

3:15PM **Summary of Discussion on Issue #2**
*Christopher Frey, Workshop Chair*

- Writing Assignments/Session

4:15PM **Observer Comments**

4:45PM **Closing Remarks**

5:00PM A D J O U R N

# APPENDIX D

# WORKSHOP CHARGE

# Workshop on Selecting Input Distributions for Probabilistic Assessment

U.S. Environmental Protection Agency
New York, NY
April 21-22, 1998

## Charge to Experts/Discussion Issues

This workshop is being held to discuss issues associated with the selection of probability distributions to represent exposure factors in a probabilistic risk assessment. The workshop discussions will focus on generic technical issues applicable to any exposure data. It is not the intent of this workshop to formulate decisions specific to any particular exposure factors. Rather, the goal of the workshop is to capture a discussion of generic issues that will be informative to Agency assessors working with a variety of exposure data.

On May 15, 1997, the U.S. Environmental Protection Agency (EPA) Deputy Administrator signed the Agency's "Policy for Use of Probabilistic Analysis in Risk Assessment." This policy establishes the Agency's position that "such probabilistic analysis techniques as Monte Carlo Analysis, given adequate supporting data and credible assumptions, can be viable statistical tools for analyzing variability and uncertainty in risk assessments." The policy also identifies several implementation activities designed to assist Agency assessors with their review and preparation of probabilistic assessments. These activities include a commitment by the EPA Risk Assessment Forum (RAF) to organize workshops or colloquia to facilitate the development of distributions for exposure factors.

In the summer of 1997, a technical panel, convened under the auspices of the RAF, began work on a framework for selecting input distributions for use in Monte Carlo analyses. The framework emphasized parametric methods and was organized around three fundamental activities: selecting candidate theoretical distributions, estimating the parameters of the candidate distributions, and evaluating the quality of the fit of the candidate distributions. In September of 1997, input on the framework was sought from a 12 member panel of experts from outside of the EPA. The recommendations of this panel include:

- expanding the framework's discussion of exploratory data analysis and graphical methods for assessing the quality of fit,
- discussing distinctions between variability and uncertainty and their implications,
- discussing empirical distributions and bootstrapping,
- discussing correlation and its implications,
- making the framework available to the risk assessment community as soon as possible.

Subsequent to receiving this input, some changes were made to the framework and it was applied to selecting distributions for three exposure factors: water intake per body weight, inhalation rate, and residence time. The results of this work are presented in the attached report entitled "Development of Statistical Distributions for Exposure Factors."

Applying the framework to the three exposure factors highlighted several issues. These issues resolved into two broad categories: issues associated with the representativeness of the data, and issues associated with using the empirical distribution function (or resampling techniques) versus using a theoretical parametric distribution function. Summaries for these issues are presented in the attached issue papers. These issues will be the focal point for discussions during this workshop. The following questions are intended to help structure and guide these discussions. In addressing these questions, workshop participants are asked to consider: what do we know today that can be applied to answering the question or providing additional guidance on the topic; what short term studies (e.g., numerical experiments) could be conducted to answer the question or provide additional guidance; and what longer term research may be needed to answer the question or provide additional guidance.

## Representativeness (Issues Paper #1)

1) The Issue Paper

Checklists I through IV in the issue paper present a framework for characterizing and evaluating the representativeness of exposure data. This framework is organized into three broad sets of questions: questions related to differences in populations, questions related to differences in spatial coverage and scale, and questions related to differences in temporal scale. Do these issues cover the most important considerations for representativeness? Are the lists of questions associated with each issue complete? If not, what questions should be added?

In a tiered approach to risk assessment (e.g., a progression from simpler screening level assessments to more complex assessments), how might the framework be tailored to each tier? For example, is there a subset of questions that adequately addresses our concerns about representativeness for a screening level risk assessment?

2) Sensitivity

The framework asks how important are (or how sensitive is the analysis to) population, spatial, and temporal differences between the sample (for which you have the data) and the population of interest. For example, to what extent do these differences affect our estimates of the mean and variance of the population and what is the magnitude and direction of these effects?

What guidance can be provided to help answer these questions? What sources of information exist to help with these questions? Having answered these questions what are the implications for the use of the data (e.g., use of the data may be restricted to screening level assessments in certain

circumstances)? What differences could be considered critical (i.e., what differences could lead to the conclusion that the assessment can't be done without the collection of additional information)?

3) Adjustments

The framework asks, is there a reasonable way of adjusting or extrapolating from the sample (for which you have data) to the population of interest in terms of the population, spatial, and temporal characteristics? If so, what methods should be used? Is there adequate information available to implement these methods?

What guidance can be provided to help answer these questions? Can exemplary methods for making adjustments be proposed? What sources of information exist to help with these questions? What research could address some of these issues?

Section 5 of the issue paper on representativeness describes methods for adjustments to account for differences in population and temporal scales. What other methods exist? What methods are available for spatial scales? Are there short-term studies that can be done to develop these methods further? Are there data available to develop these methods further? Are there numerical experiments (e.g., simulations) that can be done to explore these methods further?

## Empirical Distribution Functions and Resampling Versus Parametric Distributions (Issues Paper #2)

1) Selecting the EDF or PDF

What are the primary considerations for assessors in choosing between the use of theoretical parametric distribution functions (PDFs) and empirical distribution functions (EDFs) to represent an exposure factor? Do the advantages of one method significantly outweigh the advantages of the other? Is the choice inherently one of preference? Are there situations in which one method is clearly preferred over the other? Are there circumstances in which either method of representation should not be used?

2) Goodness of Fit

On what basis should it be decided whether or not a data set is adequately represented by a fitted analytic distribution? What role should the goodness-of-fit test statistic play (e.g., chi-square, Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises, etc.)? How should the level of significance, i.e., p-value, of the goodness of fit statistic be chosen? What are the implications or consequences for exposure assessors when acceptance/rejection is dependent on the goodness of fit statistic chosen and an arbitrary level of statistical significance? What role should graphical examination of the quality of fit play in the decision as to whether a fit is acceptable or not?

When the only data readily available are summary statistics (e.g., selected percentiles, mean, and variance), are fits to analytic distributions based on those summary statistics acceptable? Should any limitations or restrictions be placed in these situations?

When the better known theoretical distributions (e.g., lognormal, gamma, Weibull, log-logistic, etc.) cannot provide an acceptable fit to a particular set of data, is there value in testing the fit of the more flexible generalized distributions (e.g., the generalized gamma and generalized F distributions) even though they are considerably more complicated and difficult to work with?

3) Uncertainty

Are there preferred methods for assessing uncertainty in the fitted parameters (e.g., methods based on maximum likelihood and asymptotic normality, bootstrapping, etc.)?

# APPENDIX E

# BREAKOUT SESSION NOTES

# APPENDIX E

## SMALL GROUP DISCUSSIONS/BRAINWRITING SESSIONS

During the workshop, the experts worked at times in smaller groups to discuss specific technical questions. Some of these sessions involved open discussions. Other sessions involved "brainwriting," during which individuals captured their thoughts on paper, in sequence, and then discussed similar and/or opposing views within each group. The outcomes of these sessions were captured by group rapporteurs and individual group members and are summarized below. This summary represents a transcription of handwritten notes and are, as such, considered rough working notes. Information from these smaller group discussions was presented and deliberated in the plenary session, and partially forms the basis of the points presented in the main text of this report.

***What information is required to fully specify a problem definition?***

- # Population at risk
- # Sample under study (include biases)
- # Spatial extent of exposure—micro, meso, macro scale
- # Exposure-dose relationship
- # Dose-response–risk relationship
- # Temporal extent (hours, days, months, years)
- # Temporal variability about trend
- # What is the "acceptable error"?
  - — yes/no
  - — categorization
  - — continuous
  - — quantitative
- # Variability/uncertainty partitioning
  - — not needed
  - — desirable
  - — mandatory
- # User of output
  - — scientific community
  - — regulatory community
  - — general public

One expert noted that the "previous problem definition" forces the blurring of the boundaries between modeling and problem description—for example, many may not consider the dose-exposure–risk relationship to be part of the problem definition.

Another expert asked, "How much information do we have to translate from measured value to population of concern?" He described the population of concern, surrogate population, individuals sampled from the surrogate population, and how well measured value represents true value. Another agreed, emphasizing the importance of temporal, spatial, and temporal-spatial representativeness (e.g., Idaho potatoes versus Maine potatoes).

Other issues in problem definition include:

- In the context of environmental remediation, a problem is defined in terms of what level of residual risk can be left on the site. The degree of representativeness needed is dependent on the land use scenario.

  Several alternative scenarios of future land use, population, etc. might be defined and analyzed. Problem definition might include establishing budget limits (for assessment and remediation); this might dictate limits on future land use and the need for evaluation.

- A problem needs to be specified in space (location), time (over what duration), and whom (person or unit). Some of these definitions may be concrete (e.g., in terms of spatial locations around a site) while some may be more vague, such as persons who live on a brownfield site (which may change over time with mobility, new land use, etc.). The problem addresses a future context, and must therefore be linked to observable data by a model/set of assumptions. The problem definition should include these models (no population change over time) or assumptions (exposure calculated over 50- year duration/time frame).

- One must define the health outcome being targeted (e.g., acute vs. cancer vs. developmental).

  Define how you will link the exposure measure to a model for hazard and/or risk (margin of exposure has different data needs from an estimate of population risk). Also, one should consider the *type* of observation being evaluated (blood measurements vs. dietary vs. ecological). This is more likely to have an impact on the representativeness of the data sample than anything else.

  Define the target risk level; this will dictate what kind of data will be necessary.

  Another panelist agreed these are important points but questioned, however, whether these factors were part of problem definition.

- Specify the scope and purpose of the assessment (e.g., regulatory decision, set cleanup standards, etc.)

- Determining how much error we are willing to live with will determine how representative the data are.

- Specify the population of concern (who they are, where they live, what kinds of activities they are involved with).

- Problem definition is the most critical part of the process, and all stakeholders should be involved as much as possible. If the stakeholders come to a common understanding of the objectives of the process, the situation becomes focused.

- Although EPA has provided much guidance for problem definition (DQOs, DQAs, etc.), what data are necessary (and to what extent it must be representative) is a function of each individual problem. Certain basic questions are common to all problem definitions (who, what, when, how); the degree to which each basic question is important is a function of the actual problem/situation.

Decision performance requirements: What is acceptable at a specific site for a specific problem (i.e., what is the degree of decision error)? An answer to this question should be decided up front as much as possible to alleviate "bias" concerns.

# Attributes of the exposed population are key issues:

— Who are they?
— What are their activities/behaviors?
— Where are they?
— When do they engage in activities and for how long?
— Why are certain activities performed?

# The potential imprecision of "national" populations seems significant. Scale is important; maybe regional is as large as it gets.

# If representativeness is a property of the population, then we should focus on methods for collecting more specific data.

# Variability within a super-population (e.g., a national study) provides useful, quantifiable bounds to potential bias and gives an upper bound on the variability that could be found in a subpopulation. This suggests that there are quantitative ways to guide the use "reduce sparingly."

# The assessor needs to ask the following questions: Is a risk assessment necessary? What is the level of detail needed for the decision at hand? What is the scope of the problem? For example,

— Who is at risk?
— Who has standing [e.g., stakeholders]?
— Who has special concerns?
— What is of concern?
— When are people exposed? (timeframe [frequency and duration], chronic vs. acute, level of time steps needed)
— Where are people exposed—spatial considerations; scope of the problem (national, regional, site?)
— How are people exposed?

# The time step used in the model must be specified. The assessor must distinguish between distribution needed for a one-day time step as compared to a one-year time step. Some models may run at different time steps (e.g., drinking water at a one-week time step to include seasonal variation; body weight at a one-year time step to include growth of a child.)

# Consideration of a tiered approach is important in problem formulation. How are data to be used? If data are to be used in a screening manner, then conservativeness is even more important than representativeness. If more than a screening assessment is proposed, the assessor should consider *what is the value added* from more complex analyses (site-specific data collection, modeling, etc.).

- As probabilistic methods continue to be developed, it will become increasingly important to specify constraints in distribution. Boundaries exist. For example, no person can eat multiple food groups at the 95th percentile.

- Two panelists noted that tiered approaches would not change the problem definition. Generally, the problem is: Under an agreed set of exposure conditions, will the population of concern experience unacceptable risks? This question would not change with a more or less sophisticated (tiered) assessment.

- When evaluating unknown future population characteristics, we are dealing with essentially unknown conditions. It is not feasible, therefore, to have as a criterion that additional information will not significantly change the outcome of the analysis. Instead, the problem needs to be defined in terms of a precise definition of population (in time and space) which is to be protected. To the extent that this is uncertain, it needs to be defined in a generalized, generic manner.

- Considerations of the "external" representativeness of the data to the population of concern is absolutely critical for "on the ground" risk assessments. The "internal" validity of the data is often a statistical question. It seems more important to ensure that the outcome of the assessment will not change based on the consideration of "external" representativeness of the data set to the population of concern.

***What constitutes (lack of) representativeness?***

**General**

The issue of data representativeness begs the question "representative of what?" In many (most?) cases, we are working backwards, using data in hand for purposes that may or may not be directly related to the reason the data were collected in the first place. Ideally, we would have a well-posed assessment problem with well-defined assessment endpoints. From that starting point, we would collect the relevant data necessary for good statistical characterization of the key exposure factors.

More generally, we are faced with the question, "Can I use these data in my analysis?" To make that judgment fairly, we would have to go through a series of questions related to the data itself and to the use we intend to make of the data. We usually ignore many of these questions, either explicitly or implicitly. The following is an attempt at listing the issues that ought to affect our judgment of data relevance.

**Sources of Variability and Uncertainty Related to the Assessment of Data Representativeness**

EPA policy sets the standard that risk assessors should seek to characterize central tendency and plausible upper bounds on both individual risk and population risk for the overall target population as well as for sensitive subpopulations. To this extent, data representativeness cannot be separated from the assessment endpoint(s). The following outlines some of the key elements affecting data representativeness. The elements are not mutually exclusive.

**Exposed Population**

general target population
particular ethnic group
known sensitive subgroup (children, elderly, asthmatics, etc.)

occupational group (applicators, etc.)
age group (infant, child, teen, adult, whole life)
sex
activity group (sport fishermen, subsistence fishermen, etc.)

**Geographic Scale, Location**

trends (stationary, non-stationary behaviors)
past, present, future exposures
lifetime exposures
less-than-lifetime exposures (hourly, daily, weekly, annually, etc.)
temporal characteristics of source(s), continuous, intermittent, periodic, concentrated (spike), random

**Exposure Route**

inhalation
ingestion (direct, indirect)
dermal (direct) contact (by activity, e.g., swimming)
multiple pathways

**Exposure/Risk Assessment Endpoint**

cancer risk
non-cancer risk (margin of exposure, hazard index)
potential dose, applied dose, internal dose, biologically effective dose
risk statistic
mean, uncertainty percentile of mean
percentile of a distribution (e.g., 95th percentile risk)
uncertainty percentile of variability percentile (upper credibility limit on 95th percentile risk)
plausible worst case, uncertainty percentile of plausible worst case

**Data Quality Issues**

direct measurement, indirect measurement (surrogates)
modeling uncertainties
measurement error (accuracy, precision, bias)
sampling error (sample size, non-randomness, independence)
monitoring issues (short-term, long-term, stationary, mobile)

**#** Almost all data used in risk assessment is not representative in one or more ways. What is important is the effect the lack of representativeness has on the risk assessment in question. If the water pathway, for example, is of minor concern, it will not matter if the water-consumption rate distribution is not representative.

A lack of representativeness could mean the risk assessment results fail to be protective of public health or grossly overestimate risks.

The Issue Paper is helpful in describing the ways in which distributions can be nonrepresentative. It can guide the selection of the input distributions.

# Representativeness needs to be considered in the context of the decision performance requirements. Factors that could have a major impact in terms of one problem/site need not have the same impact across all problems/sites. Decision performance requirements should therefore be considered with problem-site-specific goals and objectives factored into the process.

# The definition of representativeness depends on how much error we are willing to live with. What is "good enough" will be case specific. Going through some case studies using assessments done for different purposes can shed some light on defining representativeness. "With regard to exposure factors, we [EPA] need to do a better job at specifying or providing better guidance on how to use the data that are available." For example, the soil ingestion data for children are limited, but may be good enough to provide an estimate of a mean. The data are not good enough to support a distribution or a good estimate of a high-end value.

# Representativeness measures the degree to which a sample of values for a given endpoint accurately and precisely (*adequately*) describes the value(s) of that endpoint likely to be seen in a target population.

# A number of issues relate to the lack of representativeness which one can use to decide upon use of a sample in a given case: The context of the observation is important. In addition to those mentioned in the Issues Paper (demographic, technical, social), other concerns include what is being measured: environmental sample (water, air, soil) versus human recall (diet) versus tissue samples in humans (e.g., blood). In most cases, provided good demographic and social information is available on key issues associated with the exposure, adjustment can be made to make a sample representative for a new population. Technical issues sometimes must be "guessed" from one sample to another (key issues like different or poor analytic techniques, altered consumption rates, etc.).

# A sample should not be used if it is flawed due to one of the following factors:

1) inappropriate methods (sample design and technical methods)
2) lack of descriptors (demographic, technical, social) to make adjustments
3) inadequate size for target measure

The above applies to the internal analysis of a sample. Human recall includes behavioral activities (e.g., time spent outdoors or indoors, number of days away from site).

# Identifying differences (as defined by the final objective) between characteristics of the subject population and the surrogate population will generally be subjective because there is usually no data for the subject population. Differences might be due to socioeconomic differences, race, or climate. Lack of representativeness should not be "too rigid" partly due to uncertainties and partly because the subject population usually includes a future population that is even less well defined than the current population.

The surrogate population may overlap (as in age/sex distribution) with the target population. A context is needed to determine what constitutes "lack of representativeness." For example, if soil ingestion is not related to gender, then while the surrogate population may be all female, it may not imply that the estimates from the surrogate population cannot be used for a target population

(including males and females). Bottom line: the factor being represented (such as gender) needs to be related to the outcome (soil ingestion) before the non-representativeness is important. Lack of representativeness "depends" in this sense on the association.

Another panelist expanded on the above, noting that the outcome determines the representativeness of the surrogate data set. If in the eyes of the "beholder" the data are "equivalent" they represent the actual population well. Defining representativeness is like defining art. One cannot describe it well; it is easily recognized but recognition is observer-dependent. We should strive to remove subjectivity as best as possible without making inflexible choices.

**#** Representativeness suggests that our exposure/risk model results are a reasonable approximation of reality. At minimum, they pass a straight-face test. Representativeness could therefore be assessed via model calibrations and validation.

**#** Representativeness often cannot be addressed unless an expert-judgment-based approach is used. It requires brainstorming based upon some knowledge of how the target population may differ from the surrogate one. In the long run, collection of more data is needed to reduce the non-representativeness of those distributions upon which decisions are based.

**#** Define the characteristics to be examined, define the population to be evaluated, select a statistically significant sample that reflects defined characteristics of the population (another expert noted that statistical significance has little relevance to the problem of representativeness—the issue is the degree of uncertainty or bias). Ensure randomness of a sample to capture the entire range of population characteristics. (Another noted that the problem is that we usually don't have such a sample but have to make a decision or take action now. If we can quantitatively evaluate representativeness, then we can at least make objective determinations of whether this lack of representativeness will materially affect the decisions.)

**#** The degree of bias that exists between a data set or sample and the problem at hand—is the sample even relevant to the problem? Types:

Scenario: Is a "future residential" scenario appropriate to the problem at hand?

Model: Is a multiplicative, independent-variable model appropriate?

Variables: Is a particular study appropriate to the problem? Is it biased? Uncertain?

**#** Two experts agreed that statistical significance has little relevance to the problem of representativeness. A well-designed controlled randomized study yielding two results can be "representative" of the mean and dispersion, albeit highly imprecise.

**#** Representativeness exists when the data sample is drawn at random from the population (including temporal and spatial characteristics) of concern, or is a census in the absence of measurement error. This condition is potentially lacking when using surrogate data that are for a population that differs in any way from the population of concern. Important differences include:

— characteristics of individuals (e.g., age, sex, etc.)

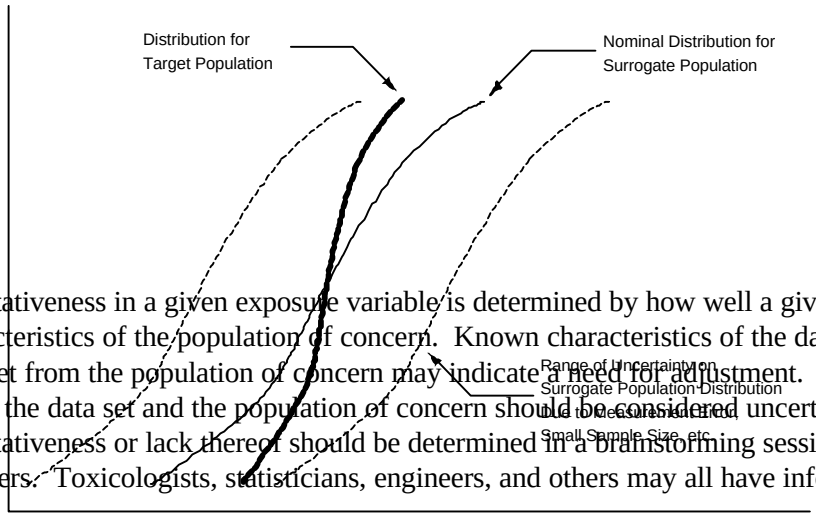
— geographic locations
— averaging time
— dynamics of population characteristics over the time frame needed in the study
— large measurement errors

Non-representativeness poses a problem if we have biases in any statistical interest (i.e., lack of representativeness can lead to biases in the mean, standard deviation, 95th percentile, etc).

Bias, or lack of accuracy, is typically more important than lack of precision. For example, we can expect some imprecision in our estimate of the 95th percentile of a population characteristic (e.g., intake rate) due to lack of relevant "census" data, but we hope that on average our assessment methods do not produce a bias or systematic error.

Conversely, if we have a large amount of uncertainty in our estimates for a sample distribution, then it is harder to claim non-representativeness than when a particular distribution for a surrogate is estimated.

In the following example, the distribution for the surrogate population is non-representative of the target population since it has too wide a variance. However, the uncertainty in the surrogate encompasses outcomes which could include the target population. Thus, in this case it may be difficult to conclude, based upon the wide range of uncertainty, that the surrogate is non-representative.




**#** Representativeness in a given exposure variable is determined by how well a given data set reflects the characteristics of the population of concern. Known characteristics of the data that distinguish the data set from the population of concern may indicate a need for adjustment. Areas of ignorance regarding the data set and the population of concern should be considered uncertainties. Representativeness or lack thereof should be determined in a brainstorming session among stakeholders. Toxicologists, statisticians, engineers, and others may all have information that

bears on the representativeness of the data. Known or suspected difference between the data set and the population of concern diminish representativeness.

# The question as to what constitutes representativeness is contingent on the problem definition—that is, who is to be represented, at what point in time, etc. If the goal is to represent a well-characterized population in the present, representativeness for a given parameter (e.g., drinking water consumption) should be evaluated based on the match of the surrogate data to the data for the population of concern relative to key correlates of the parameter (e.g., for drinking water volume, age, average ambient temperature, etc.). If, on the other hand, the population of concern is not well characterized in the present, or if the intent of the risk assessment is to address risk into the indefinite future, representativeness does not appear to have a clear meaning. The goal in such cases should be to define reasonable screening characteristics of a population at an indefinite point in time (e.g., maximum value, minimum value, estimated 10th percentile, estimated 90th percentile) and select such values from a semi-quantitative analysis of the available surrogate data.

# A representative surrogate sample is one that adds information to the assessment beyond the current state of knowledge. However, both the degree to which it adds information and the remaining uncertainty in the risk characterization must be identified.

# Suggestion: Replace the word representative with "useful and informative."

# A data set is representative of a characteristic of the population if it can be shown that differences between the data set and the population of concern will not change the outcome of the assessment. In practice, a data set should be considered in terms of its similarity and difference to the population of concern and expectations as to how the differences might change the outcome. Of course, these expectations may lead to adjustments in the data set which would make it potentially more representative of the population.

# In part, what degree of comfort the risk assessor/reviewer needs to have for the population under consideration determines how representative data have to be. Also of concern is where in the population of concern observations will take place. Are we comparing data mean or tails (outliers)? What degree of uncertainty and variability between the population of concern and the surrogate data is the assessor willing to live with?

# We may be using the term "representativeness" too broadly. Many of the issues seem to address the "validity" of the study being evaluated. However, keeping with the broad definition, the following apply to internal representativeness:

— *Measurement reliability*. Measurement reliability refers whether the study correctly measures what it set out to measure and provides some basis for evaluating the error in measurement.

— *Bias in sampling.* Bias in sampling presupposes that there is a "population" that was sampled and not just a haphazard collection of observations and measurements.

— *Statistical sampling error.*

The following issues apply to external representativeness:

— Did the study measure what we need to know (e.g., short-term vs. long-term studies). If there is a statistical procedure for translating measurements into an estimate of the needed values, the validity and errors involved must be considered.

— "Representativeness" implies that the sample data is appropriate to another population in an assessment.

***What considerations should be included in, added to, or excluded from the checklists?***

**#** Expand to include other populations of concern (e.g., ecological, produce). The issue paper and checklist seem to presuppose that the population of concern is the human population.

**#** Include more discussion on criteria for determining if question is adequately and appropriately answered.

**#** Clarify definitions (e.g., internal versus external)

**#** Include "worked" examples:

— Superfund-type risk assessment
— Source-exposure-dose-effect-risk example
— Include effect of bias, misclassification, and other problems

**#** Ask if factors are known or suspected of being associated with the outcome measured? Was the distribution of factors known or suspected to be associated with the outcome spanned by the sample data? Focus on outcome of risk assessments (if populations are different, does it make any real difference in the outcome of the assessment?).

**#** How will the exposures be used in risk assessment? For example, is the sample representative enough to bound the risk?

**#** In judging the quality of a sample, especially with questionnaire-based data, determine whether a consistency check was put in the forms and the degree to which individual samples are consistent. Risk assessors must be able to review the survey instrument.

**#** Internal and external lists may each need some reorganization (for example, measurement issues vs. statistical bias and sampling issues for "internal;" extrapolation to a different population vs. reanalysis/reinterpretation of measurement data for "external").

**#** Is a good set of subject descriptors (covariates such as age, ethnicity, income, education, or other factors that can affect behavior or response) available for both the population sampled and population of concern to allow for correlations and adjustments based on these?

- How valuable would some new or additional data collection be for the population of concern to confirm the degree of representativeness of the surrogate population and better identify and estimate the adjustment procedure?

- What is the endpoint of concern and what decision will be based on the information that is gathered? Since risk assessment involves a tiered approach, checklist should focus around the following type of question: Do I have enough information about population (type, space, time) that allow answering the questions at this tier and is my information complete enough that I can make a management decision? Do I need to go through all of the checklists before I can stop? (Questioning application/implementation)

- The checklists should address how much is known about the population of concern relative to the adaptation of the surrogate data. If the population of concern is inadequately characterized, then the ability to consider the representativeness of the surrogate data is limited, and meaningless adjustment will result.

- One consideration that is missing from the checklists is the fact that risk assessments are done for a variety of purposes. A screening level assessment may not need the level of detail that the checklists include. The checklists should be kept as simple and short as possible, trying to avoid redundancy.

- The checklist should be flexible enough to cover a variety of different problems and should be only a guide on how to approach the problem. The more considerations included the better.

- Guidance is needed on how to address overlap of the checklists. For example, when overlap exists (e.g., in some spatial and temporal characteristics), which questions in the checklist are critical? The guidance could use real life case studies to help focus the risk assessor on the issues that are critical to representativeness.

- Move from a linear checklist format to a flowchart/framework centered around the "critical" elements of representativeness.

- Fold in nature of tiered analysis. The requirements of a screening level assessment must be different from those of a full-blown risk assessment.

- Identify threshold (make or break) issues to the extent possible (i.e., minimum requirements).

- When biases due to lack of representativeness are suspected, how can we judge which direction those biases take (high or low?).
- Include a "box" describing cases when "nonrepresentative" and "inadequate" will need to be used in a risk assessment (which is common)....Figure 1?

- Define ambiguous terms, such as "reasonable" and "important."

- Make checklist more than binary (yes, no)—allow for qualitative evaluation of data.

Key questions: Can data be used at all? If so, do we have a great deal of confidence in it or not? Is data biased high or low? Can data be used in a quantitative, semi-quantitative, or only a qualitative manner? Standards according to which checklist items are evaluated should be consistent with stated objective (e.g., a screening assessment will require less stringent evaluation of data set than a site assessment where community concerns or economic costs are critical issues).

- # Allow for professional judgement and expert elicitation.

- # What are the representativeness decision criteria? Data only have to be good enough for the problem at hand; there are no perfect data. List some considerations pertaining to the acceptance/rejection criteria.

- # The 95th percentile of each input distribution is not needed to forecast risk at the 95th percentile with high accuracy and low uncertainty.

- # What is the study population doing? (i.e., were the sample population and study population engaged in similar activities?) Consider how their behavior affects ability to represent.

- # Combine Checklists II, III, and IV into one.

- # Distinguish between marginal distributions vs. joint distributions vs. functional relationships.

- # Distinguish variability from uncertainty. Add a crisp definition of each (e.g., Burmaster's premeeting comments).

- # Add explicit encouragement and positive incentives to collect and analyze new data.

- # Add an explicit statement that the agency encourages the development and use of new methods and that nothing in this guidance should be interpreted as blocking the use of alternative or new methods.

- # Add an explicit statement that it is always appropriate to combine information from several studies to develop a distribution for an exposure factor. (This also applies to toxicology and the development of distributions for reference doses and cancer slope factors.)

***How can one perform a sensitivity analysis to evaluate the implications of non-representativeness? How do we assess the importance of non-representativeness?***

- # The assessor should ask, "under a range of plausible adjustments from the surrogate population to the population of concern, does (or can) the risk management decision change?" That is, do these particular assumptions and their uncertainty matter? (among all others)

  Representativeness is often not that important, because risk management decisions are usually not designed to protect just the current population at a particular location, but a range of possible target populations (e.g., future site or product users) under different possible scenarios.

# Theoretically, we can come up with a "perfect" risk assessment in terms of representativeness, but if the factor(s) being evaluated is not important, then the utility of this perfectly representative data is limited. The important question to ask is: If one is wrong, what are the consequences, and what difference do the decision errors make in the estimate of the parameter being evaluated?

The question of data representativeness can be asked absent the context/model/parameter or it can be asked in the context of a decision or analysis (are the data adequate?).

The key is placing bounds on the use of the data. Assessments should be put in context and the level at which surrogate data may be representative. It should be defined in the context of the purpose of the original study. Two other factors are critical: sensitivity and cost/resource allocation. The question, therefore, is situation-specific.

# A sensitivity analysis can be conducted in the context of the following tiered approach. The importance of a parameter (as evidenced by a sensitivity analysis) is determined first, making the representativeness or non-representativeness of the non-sensitive parameters unimportant.

# Representativeness is not a standard statistical term. Statistical terms that may be preferable include bias and consistency.

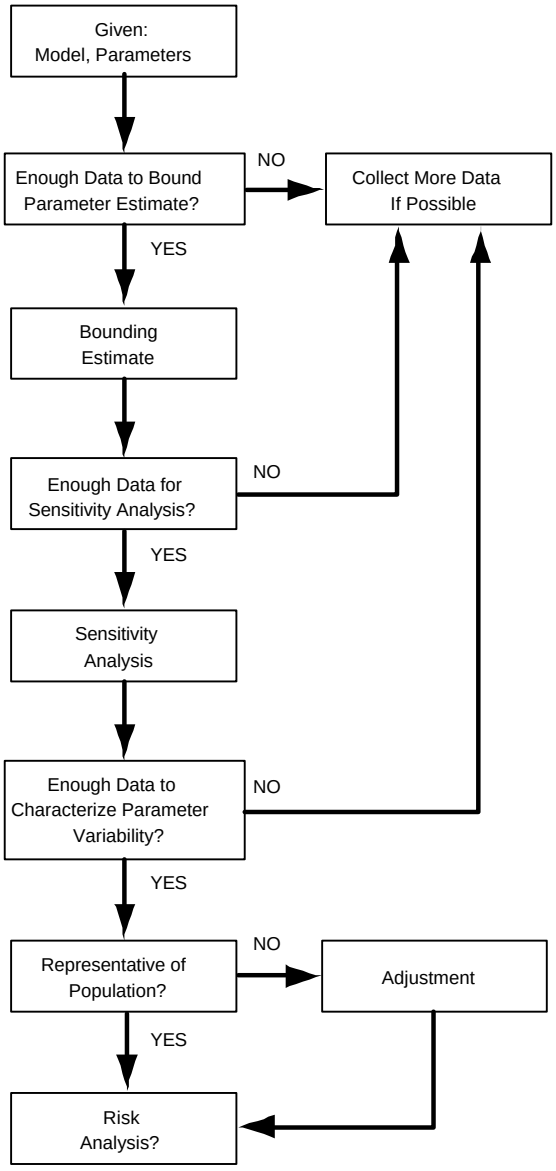
When evaluating the importance of non-representativeness, one needs to evaluate the uncertainty on the data set and on the individual. At the first level the assessor may choose a value biased high (could be a point value or a distribution that is shifted up). At the second level, can use an average, but must still be sensitive to whether acute or chronic effects are being evaluated. When looking at the individual sample it is more important to have a representative sample because the relevant data are in the tails (more important for acute toxicity). When using a mixture, representativeness is less of a problem.

***Adjustments***

# Take more human tissue samples to back calculate—this makes local population happier. Determine the need for cleanup based on tissue sample findings.

# Re-do large samples (e.g., food consumptions, tapwater consumption).

# Look at demographics, etc. and determine the most sensitive factor(s).




**#** Use a general model. Discuss with stakeholders the degree of inclusion in general. Adjust the model with survey data if it is not applicable to stakeholder. Use a special model for subpopulations if necessary.

**#** "Change of support" analysis; time-series analysis — non-CERCLA, important to the Food Quality Protection Act/

**#** Conduct three-day surveys with year-long adjustments.

**#** Hypothesis methods will work, but need to be tested.

# The group recommended holding a workshop for experts in related fields to share *existing* theory and methods on adjustment (across fields).

# General guidelines for adjustments will be acceptable, but often site-specific needs dictate what adjustments must be made.

# Example adjustment:

Fish consumption: If you collect data 3 days per week, you may miss those who might eat less—a case of inter- versus intra-individual variability.

# Adjustment is often difficult because of site specifics and evaluator bias or professional judgement.

# Sometimes it is not possible to adjust. Using an alternate surrogate data set makes it possible to set some plausible bounds to perform a screening risk assessment.

# Stratify data to see if any correlation exists.

# Start with brainstorming.

# Regression relationship versus threshold.

# Covariance; good statistical power to sample population.

# Correlation is equivalent to regression analysis as long as you keep the residual (Bayesian presentation).

# Instead of looking at the population, look at the individual (e.g., breathing rates or body weight for individuals from ages 0 to 30) to establish correlations.

# What if the population was misrepresented? For example, population of concern is sport fishermen but the national data represent other types of fishermen.

Set up a hierarchy:

- do nothing (may fall out when bounded)
- conservative/plausible upper bound
- use simple model to adjust the data (may be worth the effort if credibility issues are dealt with)
- resample/collect more data

Before considering a bounding approach (model development), consider if refining is necessary or cost/beneficial.

Are there situations in which “g-estimates” are worthwhile?

# What is gained by making adjustments?

Short-term studies overestimate variability because they do not account for interindividual variability (upper tail is overstated).

# Can we estimate the direction of biases when populations are mismatched?

If the bias is conservative, then we are being protective. But what if the bias is nonconservative (e.g., drinking water in the Mojave Desert or by construction workers)?

# Appropriate models

Simplistic:

How speculative? Identify potential damage due to credibility issues.

Complex:

Identify the bias: high (conservative); or low (different scenario used than plausible bounding analysis)?

# Unless one has a sense of the likelihood of the scenario, what does one do?

— Risk management can address it.
— Present qualitative statements about uncertainty.
— Value of information approaches (e.g., does weather change drinking water data?).

Short-term Research:

Evaluate short-term data set: make assumptions, devise models on population variability (Ryan paper) (Wallace and Buck). Look at behavior patterns, information biases. Flesh out Chris Portier's suggestion on extrapolating 3-day data to 6 months, years. This would give the assessor some confidence in extrapolating for interindividual variability.

Long-term Research:

Collect more data. Possible ORD funding? Look at breathing rates, soil ingestion, infrequently consumed items, frequently consumed items.